

## (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
18 October 2001 (18.10.2001)

PCT

(10) International Publication Number  
**WO 01/77340 A1**

- (51) International Patent Classification<sup>7</sup>: C12N 15/55, 15/62, 9/16, C07K 16/40, C12Q 1/68, 1/42, G01N 33/53 (74) Common Representative: MERCK PATENT GMBH; Frankfurter Strasse 250, 64293 Darmstadt (DE).
- (21) International Application Number: PCT/EP01/03966 (81) Designated States (*national*): CA, JP, US.
- (22) International Filing Date: 6 April 2001 (06.04.2001) (84) Designated States (*regional*): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 00107143.0 10 April 2000 (10.04.2000) EP
- (71) Applicant (*for all designated States except US*): MERCK PATENT GMBH [DE/DE]; Frankfurter Strasse 250, 64293 Darmstadt (DE).
- (72) Inventor; and
- (75) Inventor/Applicant (*for US only*): DÜCKER, Klaus [DE/DE]; Ettesterstrasse 5, 64291 Darmstadt (DE).

## Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/77340 A1

(54) Title: IDENTIFICATION OF A DUAL SPECIFICITY PHOSPHATASE: DUSP-10

(57) Abstract: DUSP10 polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing DUSP10 polypeptides and polynucleotides in diagnostic assays.

## IDENTIFICATION OF A DUAL SPECIFICITY PHOSPHATASE: DUSP-10

**Field of the Invention**

5 This invention relates to newly identified polypeptides and polynucleotides encoding such polypeptides sometimes hereinafter referred to as „dual specificity phosphatase (DUSP10)“, to their use in diagnosis and in identifying compounds that may be agonists, antagonists that are potentially useful in therapy, and to production of such polypeptides and polynucleotides.

**Background of the Invention**

10 The drug discovery process is currently undergoing a fundamental revolution as it embraces "functional genomics", that is, high throughput genome- or gene-based biology. This approach as a means to identify genes and gene products as therapeutic targets is rapidly superceding earlier approaches based on "positional cloning". A phenotype, that is a biological function or genetic disease, would be identified and this would then be tracked back to the responsible gene, based on its genetic map position.

15 Functional genomics relies heavily on high-throughput DNA sequencing technologies and the various tools of bioinformatics to identify gene sequences of potential interest from the many molecular biology databases now available. There is a continuing need to identify and characterise further genes and their related polypeptides/proteins, as targets for drug discovery.

**Summary of the Invention**

20 The present invention relates to dual specificity phosphatase-10 (DUSP10), in particular DUSP10 polypeptides and DUSP10 polynucleotides, recombinant materials and methods for their production. Such polypeptides and polynucleotides are of interest in relation to methods

25

30

of treatment of certain diseases, including, but not limited to, cancer (including but not limited to leukemia, colon carcinoma, lung cancer, prostate cancer), metastasis of tumor cells, neo-angiogenesis, epilepsy, stroke, neuronal degeneration syndromes, alzheimer, depression, schizophrenia, cardiac myotrophies, asthma, immune disorders, autoimmune diseases, inflammatory processes (including but not limited to arthritis and bowels disease, diabetes type I), osteoporosis, diabetes and diabetes associated diseases, hereinafter referred to as " diseases of the invention". In a further aspect, the invention relates to methods for identifying agonists and antagonists (e.g., inhibitors) using the materials provided by the invention, and treating conditions associated with DUSP10 imbalance with the identified compounds. In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate DUSP10 activity or levels.

### Description of the Invention

In a first aspect, the present invention relates to DUSP10 polypeptides. Such polypeptides include:

(a) A polypeptide encoded by a polynucleotide comprising the sequence of SEQ ID NO:1;

(b) A polypeptide comprising a polypeptide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;

(c) A polypeptide comprising the polypeptide sequence of SEQ ID NO:2;

(d) A polypeptide having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;

(e) the polypeptide sequence of SEQ ID NO:2; and

(f) polypeptide having or comprising a polypeptide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polypeptide sequence of SEQ ID NO:2;

(g) fragments and variants of such polypeptides in (a) to (f).

Polypeptides of the present invention are believed to be members of the dual specificity phosphatases family of polypeptides. Growth factors, cytokines and a variety of extracellular stress stimuli activate cytosolic signalling pathways, which very often converge on members of the family of mitogen activated protein kinases (MAPK). This family consists of three major subclasses, MAPK/ERK, SAPK/JNK and p38 (Marshall, C.J., Curr. Opin. Genet. Dev. 4:82-9, 1994; Karin, M., Ann. N.Y. Acad. Sci. 851:139-46, 1998; Kyriakis, J.M. et al., Bioessays 18:567-77, 1996; Ip, Y.T. et al., Curr. Opin. Cell. Biol. 10:205-19, 1998). While kinases of the MAPK/ERK subclass are usually activated by growth factors or phorbol esters, SAPK/JNK and p38 kinases are activated by inflammatory cytokines or extracellular stress stimuli like UV irradiation or osmotic stress (Marshall, C.J., Cell 80: 179-185, 1995; Treisman, R., Curr. Opin. Cell. Biol. 8:205-215, 1996). Upon activation, MAP kinases in turn phosphorylate diverse target proteins, which can be localized in the membrane/cytosolic fraction as well as in the nucleus. Dependent on the response capacity of the signal perceiving cell a wide range of intracellular and extracellular responses has been identified to be regulated by MAPK phosphorylation, ranging from cell proliferation, neoplastic transformation (e.g. of fibroblasts), differentiation (e.g. PC12 cells), DNA repair, up to induction of apoptosis on the cellular level, and for example platelet aggregation and immune system modulatory effects on the systemic level (Saklatvala, J. et al., J. Biol. Chem. 271:6586-6589, 1996; Verheij, M. et al., Nature 380:75-79, 1996; Potapova, O. et al., J. Biol. Chem. 272:14041-14044, 1997; Porter A.C. et al., Oncogene 17:1343-1352, 1998; Cross, J.V. et al., J. Biol. Chem. 274:31150-31154, 1999; Cannons, J.L. et al., J. Immunol. 163:2990-2998, 1999).

Activity of members of the MAPK/ERK, SAPK/JNK and p38 subclasses are regulated by phosphorylation/dephosphorylation on both threonine and tyrosine residues within the motifs TEY, TPY and TGY, respectively located in the activation loop of kinase subdomain VIII. Phosphorylation/dephosphorylation is mediated by selective dual specificity kinases (Dhanasekaran, N. et al., Oncogene 17:1447-55, 1998) and dual specificity phosphatases (DUSP) (Keyse, S.M., Semin Cell. Dev. Biol. 9:143-145, 1998; Camps, M. et al., FASEB J. 14:6-16, 2000). The existence of at least 9 different DUSPs implies a complex

network regulating MAPK signalling pathways. Interestingly, these DUSPs show distinct subcellular localisation as well as substrate specificities. Therefore, to inactivate several MAPK isoforms in a given cell often requires simultaneous expression of various DUSPs.

5

Herewith we disclose the sequence of a novel human dual specificity phosphatase, DUSP-10, with highest homology to the human and mouse DUSP hVH5 and NTTP1, respectively (Martell, K.J. et al., J.Neurochem. 65:1823-1833, 1995; Theodosiou, A.M. et al., Hum. Mol. Genet. 5:675-684, 1996). Compared to hVH5/NTTP-1 DUSP-10 shows a similar domain organisation and overall length of the protein sequence. N-terminal of the DUSP-10 protein sequence is a CH2 (cdc25 homology) domain, followed by the actual catalytical phosphatase domain and a long C-terminal stretch of so far unknown function. The CH2 domain is present in most of the so far known dual specificity phosphatases and has been postulated to be involved in substrate recognition and binding (Keyes, S.M. and Ginsburg, M., Trends Biochem. 18:377-378, 1993; Kwak, S.P. et al., J. Biol. Chem. 269:3596-3604, 1994; Muda, M., et al., J. Biol. Chem. 273:9323-9329, 1998).

20

DUSP-10 is located on chromosome 12 (D12S99 – D12S358). It has been postulated that this region contains a novel, so far unknown tumor suppressor gene, which might be involved in the pathogenesis of various cancer forms like breast cancer (Spirin, K.S. et al., Cancer Res. 56:2400-2404, 1996), acute lymphoblastic leukemia (Takeuchi, S. et al., Blood 87:3368-3374, 1996) or non-small cell lung cancer (Takeuchi S., et al., Cancer Res. 56:738-740, 1996). Interestingly, the expressed sequence tag (AI424957) represents an unexpected genomic aberration form of the DUSP-10 gene (see below "Further Examples") and is derived from a cDNA library made of B-cells from a chronic lymphocytic leukemia patient, implying that DUSP-10 could represent this novel tumor suppressor gene.

30

Given the great range of biological processes in which MAPK signalling is involved, modulators of dual specificity phosphatase activities can have a profound effect on these signalling processes. .

35

The biological properties of the DUSP10 are hereinafter referred to as "biological activity of DUSP10" or "DUSP10 activity". Preferably, a polypeptide of the present invention exhibits at least one biological activity of DUSP10.

5 Polypeptides of the present invention also includes variants of the aforementioned polypeptides, including all allelic forms and splice variants. Such polypeptides vary from the reference polypeptide by insertions, deletions, and substitutions that may be conservative or non-conservative, or any combination thereof. Particularly preferred variants are those in  
10 which several, for instance from 50 to 30, from 30 to 20, from 20 to 10, from 10 to 5, from 5 to 3, from 3 to 2, from 2 to 1 or 1 amino acids are inserted, substituted, or deleted, in any combination.

Preferred fragments of polypeptides of the present invention include an isolated polypeptide comprising an amino acid sequence having at  
15 least 30, 50 or 100 contiguous amino acids from the amino acid sequence of SEQ ID NO: 2, or an isolated polypeptide comprising an amino acid sequence having at least 30, 50 or 100 contiguous amino acids truncated or deleted from the amino acid sequence of SEQ ID NO: 2. Preferred fragments are biologically active fragments that mediate the  
20 biological activity of DUSP10, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also preferred are those fragments that are antigenic or immunogenic in an animal, especially in a human.

25 Fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, these variants may be employed as intermediates for producing the full-length polypeptides of the invention. The polypeptides of the present invention may be in the form of the "mature" protein or may be a part of a larger protein such as a precursor or a fusion protein. It is  
30 often advantageous to include an additional amino acid sequence that contains secretory or leader sequences, pro-sequences, sequences that aid in purification, for instance multiple histidine residues, or an additional sequence for stability during recombinant production.

35 Polypeptides of the present invention can be prepared in any suitable manner, for instance by isolation from naturally occurring sources, from

genetically engineered host cells comprising expression systems (*vide infra*) or by chemical synthesis, using for instance automated peptide synthesisers, or a combination of such methods.. Means for preparing such polypeptides are well understood in the art.

5

In a further aspect, the present invention relates to DUSP10 polynucleotides. Such polynucleotides include:

10

(a) A polynucleotide comprising a polynucleotide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polynucleotide sequence of SEQ ID NO:1;

(b) A polynucleotide comprising the polynucleotide of SEQ ID NO:1;

(c) A polynucleotide having at least 95%, 96%, 97%, 98%, or 99% identity to the polynucleotide of SEQ ID NO:1;

(d) the polynucleotide of SEQ ID NO:1;

15

(e) polynucleotide comprising a polynucleotide sequence encoding a polypeptide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;

(f) A polynucleotide comprising a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2;

20

(g) A polynucleotide having a polynucleotide sequence encoding a polypeptide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;

(h) A polynucleotide encoding the polypeptide of SEQ ID NO:2;

25

(i) A polynucleotide having or comprising a polynucleotide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polynucleotide sequence of SEQ ID NO:1;

30

(j) A polynucleotide having or comprising a polynucleotide sequence encoding a polypeptide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polypeptide sequence of SEQ ID NO:2; and

polynucleotides that are fragments and variants of the above mentioned polynucleotides or that are complementary to above mentioned polynucleotides, over the entire length thereof.

Preferred fragments of polynucleotides of the present invention include a polynucleotide comprising an nucleotide sequence having at least 15, 30, 50 or 100 contiguous nucleotides from the sequence of SEQ ID NO: 1, or an isolated polynucleotide comprising an sequence having at least 30, 50 or 100 contiguous nucleotides truncated or deleted from the sequence of SEQ ID NO: 1.

Preferred variants of polynucleotides of the present invention include splice variants, allelic variants, and polymorphisms, including polynucleotides having one or more single nucleotide polymorphisms (SNPs).

Polynucleotides of the present invention also include polynucleotides encoding polypeptide variants that comprise the amino acid sequence of SEQ ID NO:2 and in which several, for instance from 50 to 30, from 30 to 20, from 20 to 10, from 10 to 5, from 5 to 3, from 3 to 2, from 2 to 1 or 1 amino acid residues are substituted, deleted or added, in any combination.

In a further aspect, the present invention provides polynucleotides that are RNA transcripts of the DNA sequences of the present invention. Accordingly, there is provided an RNA polynucleotide that:

(a) comprises an RNA transcript of the DNA sequence encoding the polypeptide of SEQ ID NO:2;

(b) is the RNA transcript of the DNA sequence encoding the polypeptide of SEQ ID NO:2;

(c) comprises an RNA transcript of the DNA sequence of SEQ ID NO:1; or

(d) is the RNA transcript of the DNA sequence of SEQ ID NO:1;

and RNA polynucleotides that are complementary thereto.



The polynucleotide sequence of SEQ ID NO:1 shows homology with HSU27193 (Martell, K.J. et al., J. Neurochem. 65 (4), 1823-1833 (1995)). The polynucleotide sequence of SEQ ID NO:1 is a cDNA sequence that encodes the polypeptide of SEQ ID NO:2. The polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding sequence of SEQ ID NO:1 or it may be a sequence other than SEQ ID NO:1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2. The polypeptide of the SEQ ID NO:2 is related to other proteins of the dual specificity phosphatases family, having homology and/or structural similarity with DUS8\_HUMAN (Martell, K.J. et al., J. Neurochem. 65 (4), 1823-1833 (1995)).

Preferred polypeptides and polynucleotides of the present invention are expected to have, *inter alia*, similar biological functions/properties to their homologous polypeptides and polynucleotides. Furthermore, preferred polypeptides and polynucleotides of the present invention have at least one DUSP10 activity.

Polynucleotides of the present invention may be obtained using standard cloning and screening techniques from a cDNA library derived from mRNA in cells of human blood, brain, colon, gall bladder, germ cell, heart, kidney, liver, lung, parathyroid, placenta, prostate, spleen, stomach, thymus and uterus, (see for instance, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

When polynucleotides of the present invention are used for the recombinant production of polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself, or the coding sequence for the mature polypeptide in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain

preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, Proc Natl Acad Sci USA (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Polynucleotides that are identical, or have sufficient identity to a polynucleotide sequence of SEQ ID NO:1, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification reaction (for instance, PCR). Such probes and primers may be used to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of other genes (including genes encoding paralogs from human sources and orthologs and paralogs from species other than human) that have a high sequence similarity to SEQ ID NO:1, typically at least 95% identity. Preferred probes and primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50, if not at least 100 nucleotides. Particularly preferred probes will have between 30 and 50 nucleotides. Particularly preferred primers will have between 20 and 25 nucleotides.

A polynucleotide encoding a polypeptide of the present invention, including homologs from species other than human, may be obtained by a process comprising the steps of screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof, preferably of at least 15 nucleotides; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such hybridization techniques are well known to the skilled artisan. Preferred stringent hybridization conditions include overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA; followed by washing the filters in 0.1x SSC at about 65°C. Thus the present invention also includes isolated polynucleotides, preferably with a nucleotide sequence of at least 100, obtained by screening a library under stringent hybridization conditions with a labeled

probe having the sequence of SEQ ID NO:1 or a fragment thereof, preferably of at least 15 nucleotides.

5 The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide does not extend all the way through to the 5' terminus. This is a consequence of reverse transcriptase, an enzyme with inherently low "processivity" (a measure of the ability of the enzyme to remain attached to the template during the polymerisation reaction), failing to complete a DNA copy of the mRNA template during first strand cDNA synthesis.

10 There are several methods available and well known to those skilled in the art to obtain full-length cDNAs, or extend short cDNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman et al., Proc Nat Acad Sci USA 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the  
15 Marathon (trade mark) technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon (trade mark) technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to  
20 amplify the "missing" 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using 'nested' primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific  
25 primer that anneals further 5' in the known gene sequence). The products of this reaction can then be analysed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the  
30 5' primer.

Recombinant polypeptides of the present invention may be prepared by processes well known in the art from genetically engineered host cells comprising expression systems. Accordingly, in a further aspect, the  
35 present invention relates to expression systems comprising a

polynucleotide or polynucleotides of the present invention, to host cells which are genetically engineered with such expression systems and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Polynucleotides may be introduced into host cells by methods described in many standard laboratory manuals, such as Davis et al., Basic Methods in Molecular Biology (1986) and Sambrook *et al.* (*ibid*). Preferred methods of introducing polynucleotides into host cells include, for instance, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as *Streptococci*, *Staphylococci*, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used, for instance, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector that is able to maintain, propagate or express a polynucleotide to produce a polypeptide in a host may be used. The appropriate polynucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as,

for example, those set forth in Sambrook *et al.*, (*ibid*). Appropriate secretion signals may be incorporated into the desired polypeptide to allow secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals.

If a polypeptide of the present invention is to be expressed for use in screening assays, it is generally preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide. If produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during intracellular synthesis, isolation and/or purification.

Polynucleotides of the present invention may be used as diagnostic reagents, through detecting mutations in the associated gene. Detection of a mutated form of the gene characterised by the polynucleotide of SEQ ID NO:1 in the cDNA or genomic sequence and which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered spatial or temporal expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques well known in the art.

Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or it may be amplified enzymatically by using PCR, preferably RT-PCR, or other amplification techniques prior to

analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled DUSP10 nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence difference may also be detected by alterations in the electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (see, for instance, Myers *et al.*, Science (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton *et al.*, Proc Natl Acad Sci USA (1985) 85: 4397-4401).

An array of oligonucleotide probes comprising DUSP10 polynucleotide sequence or fragments thereof can be constructed to conduct efficient screening of *e.g.*, genetic mutations. Such arrays are preferably high density arrays or grids. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability, see, for example, M.Chee *et al.*, Science, 274, 610-613 (1996) and other references cited therein.

Detection of abnormally decreased or increased levels of polypeptide or mRNA expression may also be used for diagnosing or determining susceptibility of a subject to a disease of the invention. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as a polypeptide of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Thus in another aspect, the present invention relates to a diagnostic kit comprising:

(a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO: 1, or a fragment or an RNA transcript thereof;

(b) a nucleotide sequence complementary to that of (a);

5 (c) a polypeptide of the present invention, preferably the polypeptide of SEQ ID NO:2 or a fragment thereof; or

(d) an antibody to a polypeptide of the present invention, preferably to the polypeptide of SEQ ID NO:2.

10 It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or susceptibility to a disease, particularly diseases of the invention, amongst others.

15 The polynucleotide sequences of the present invention are valuable for chromosome localisation studies. The sequence is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. 20 Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (co-inheritance of physically adjacent genes). Precise 25 human chromosomal localisations for a genomic sequence (gene fragment etc.) can be determined using Radiation Hybrid (RH) Mapping (Walter, M. Spillett, D., Thomas, P., Weissenbach, J., and Goodfellow, P., (1994) A method for constructing radiation hybrid maps of whole genomes, Nature Genetics 7, 22-28). A number of RH panels are available from Research Genetics (Huntsville, AL, USA) e.g. the GeneBridge4 RH panel (Hum Mol Genet 1996 Mar;5(3):339-46 A radiation hybrid map of the human genome. Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'Homme 30

JF, Dib C, Auffray C, Morissette J, Weissenbach J, Goodfellow PN). To determine the chromosomal location of a gene using this panel, 93 PCRs are performed using primers designed from the gene of interest on RH DNAs. Each of these DNAs contains random human genomic fragments maintained in a hamster background (human / hamster hybrid cell lines). These PCRs result in 93 scores indicating the presence or absence of the PCR product of the gene of interest. These scores are compared with scores created using PCR products from genomic sequences of known location. This comparison is conducted at <http://www.genome.wi.mit.edu/>. The gene of the present invention maps to human chromosome 20 (D20S99 - D20S358).

The polynucleotide sequences of the present invention are also valuable tools for tissue expression studies. Such studies allow the determination of expression patterns of polynucleotides of the present invention which may give an indication as to the expression patterns of the encoded polypeptides in tissues, by detecting the mRNAs that encode them. The techniques used are well known in the art and include in situ hybridisation techniques to clones arrayed on a grid, such as cDNA microarray hybridisation (Schena *et al*, Science, 270, 467-470, 1995 and Shalon *et al*, Genome Res, 6, 639-645, 1996) and nucleotide amplification techniques such as PCR. A preferred method uses the TAQMAN (Trade mark) technology available from Perkin Elmer. Results from these studies can provide an indication of the normal function of the polypeptide in the organism. In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by an alternative form of the same gene (for example, one having an alteration in polypeptide coding potential or a regulatory mutation) can provide valuable insights into the role of the polypeptides of the present invention, or that of inappropriate expression thereof in disease. Such inappropriate expression may be of a temporal, spatial or simply quantitative nature.

The polypeptides of the present invention are expressed in blood, brain, colon, gall bladder, germ cell, heart, kidney, liver, lung, parathyroid, placenta, prostate, spleen, stomach, thymus and uterus.



A further aspect of the present invention relates to antibodies. The polypeptides of the invention or their fragments, or cells expressing them, can be used as immunogens to produce antibodies that are immunospecific for polypeptides of the present invention. The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

Antibodies generated against polypeptides of the present invention may be obtained by administering the polypeptides or epitope-bearing fragments, or cells to an animal, preferably a non-human animal, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., *Nature* (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy*, 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies, such as those described in U.S. Patent No. 4,946,778, can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms, including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography. Antibodies against polypeptides of the present invention may also be employed to treat diseases of the invention, amongst others.

Polypeptides and polynucleotides of the present invention may also be used as vaccines. Accordingly, in a further aspect, the present invention relates to a method for inducing an immunological response in a mammal that comprises inoculating the mammal with a polypeptide of the present invention, adequate to produce antibody and/or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said animal from disease, whether that disease is already established within the individual or not. An immunological response in a

mammal may also be induced by a method comprises delivering a polypeptide of the present invention *via* a vector directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases of the invention. One way of administering the vector is by accelerating it into the desired cells as a coating on particles or otherwise. Such nucleic acid vector may comprise DNA, RNA, a modified nucleic acid, or a DNA/RNA hybrid. For use a vaccine, a polypeptide or a nucleic acid vector will be normally provided as a vaccine formulation (composition). The formulation may further comprise a suitable carrier. Since a polypeptide may be broken down in the stomach, it is preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions that may contain anti-oxidants, buffers, bacteriostats and solutes that render the formulation isotonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions that may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Polypeptides of the present invention have one or more biological functions that are of relevance in one or more disease states, in particular the diseases of the invention hereinbefore mentioned. It is therefore useful to identify compounds that stimulate or inhibit the function or level of the polypeptide. Accordingly, in a further aspect, the present invention provides for a method of screening compounds to identify those that stimulate or inhibit the function or level of the polypeptide. Such methods identify agonists or antagonists that may be employed for therapeutic and prophylactic purposes for such diseases of the invention as hereinbefore mentioned. Compounds may be identified from a variety of sources, for

example, cells, cell-free preparations, chemical libraries, collections of chemical compounds, and natural product mixtures. Such agonists or antagonists so-identified may be natural or modified substrates, ligands, receptors, enzymes, etc., as the case may be, of the polypeptide; a structural or functional mimetic thereof (see Coligan *et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991)) or a small molecule.

The screening method may simply measure the binding of a candidate compound to the polypeptide, or to cells or membranes bearing the polypeptide, or a fusion protein thereof, by means of a label directly or indirectly associated with the candidate compound. Alternatively, the screening method may involve measuring or detecting (qualitatively or quantitatively) the competitive binding of a candidate compound to the polypeptide against a labeled competitor (e.g. agonist or antagonist). Further, these screening methods may test whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells bearing the polypeptide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed. Further, the screening methods may simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide of the present invention, to form a mixture, measuring a DUSP10 activity in the mixture, and comparing the DUSP10 activity of the mixture to a control mixture which contains no candidate compound.

Polypeptides of the present invention may be employed in conventional low capacity screening methods and also in high-throughput screening (HTS) formats. Such HTS formats include not only the well-established use of 96- and, more recently, 384-well micotiter plates but also emerging methods such as the nanowell method described by Schullek *et al.*, Anal Biochem., 246, 20-29, (1997).

Fusion proteins, such as those made from Fc portion and DUSP10 polypeptide, as hereinbefore described, can also be used for high-throughput screening assays to identify antagonists for the polypeptide of the present invention (see D. Bennett *et al.*, J Mol Recognition, 8:52-58 (1995); and K. Johanson *et al.*, J Biol Chem, 270(16):9459-9471 (1995)).

## Screening techniques

5 The polynucleotides, polypeptides and antibodies to the polypeptide of the present invention may also be used to configure screening methods for detecting the effect of added compounds on the production of mRNA and polypeptide in cells. For example, an ELISA assay may be constructed for measuring secreted or cell associated levels of polypeptide using monoclonal and polyclonal antibodies by standard methods known in the art. This can be used to discover agents that may inhibit or enhance the production of polypeptide (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

15 A polypeptide of the present invention may be used to identify membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the polypeptide is labeled with a radioactive isotope (for instance,  $^{125}\text{I}$ ), chemically modified (for instance, biotinylated), or fused to a peptide sequence suitable for detection or purification, and incubated with a source of the putative receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. These screening methods may also be used to identify agonists and antagonists of the polypeptide that compete with the binding of the polypeptide to its receptors, if any. Standard methods for conducting such assays are well understood in the art.

25 Examples of antagonists of polypeptides of the present invention include antibodies or, in some cases, oligonucleotides or proteins that are closely related to the ligands, substrates, receptors, enzymes, etc., as the case may be, of the polypeptide, e.g., a fragment of the ligands, substrates, receptors, enzymes, etc.; or a small molecule that bind to the polypeptide of the present invention but do not elicit a response, so that the activity of the polypeptide is prevented.

30 Screening methods may also involve the use of transgenic technology and DUSP10 gene. The art of constructing transgenic animals is well

established. For example, the DUSP10 gene may be introduced through microinjection into the male pronucleus of fertilized oocytes, retroviral transfer into pre- or post-implantation embryos, or injection of genetically modified, such as by electroporation, embryonic stem cells into host blastocysts. Particularly useful transgenic animals are so-called "knock-in" animals in which an animal gene is replaced by the human equivalent within the genome of that animal. Knock-in transgenic animals are useful in the drug discovery process, for target validation, where the compound is specific for the human target. Other useful transgenic animals are so-called "knock-out" animals in which the expression of the animal ortholog of a polypeptide of the present invention and encoded by an endogenous DNA sequence in a cell is partially or completely annulled. The gene knock-out may be targeted to specific cells or tissues, may occur only in certain cells or tissues as a consequence of the limitations of the technology, or may occur in all, or substantially all, cells in the animal. Transgenic animal technology also offers a whole animal expression-cloning system in which introduced genes are expressed to give large amounts of polypeptides of the present invention

Screening kits for use in the above described methods form a further aspect of the present invention. Such screening kits comprise:

- (a) a polypeptide of the present invention;
- (b) a recombinant cell expressing a polypeptide of the present invention;
- (c) a cell membrane expressing a polypeptide of the present invention; or
- (d) an antibody to a polypeptide of the present invention;

which polypeptide is preferably that of SEQ ID NO:2.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component.

### Glossary

The following definitions are provided to facilitate understanding of certain terms used frequently hereinbefore.

"Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an

Fab or other immunoglobulin expression library.

5 "Isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein. Moreover,  
10 a polynucleotide or polypeptide that is introduced into an organism by transformation, genetic manipulation or by any other recombinant method is "isolated" even if it is still present in said organism, which organism may be living or non-living.

15 "Polynucleotide" generally refers to any polyribonucleotide (RNA) or polydeoxribonucleotide (DNA), which may be unmodified or modified RNA or DNA. "Polynucleotides" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is  
20 mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term "polynucleotide" also  
25 includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications may be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or  
30 metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

35 "Polypeptide" refers to any polypeptide comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds,

i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques that are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications may occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present to the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, biotinylation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination (see, for instance, *Proteins - Structure and Molecular Properties*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993; Wold, F., *Post-translational Protein Modifications: Perspectives and Prospects*, 1-12, in *Post-translational Covalent Modification of Proteins*, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", *Meth Enzymol*, 182, 626-646, 1990, and Rattan *et al.*, "Protein Synthesis: Post-translational Modifications and Aging", *Ann NY Acad Sci*, 663, 48-62, 1992).

"Fragment" of a polypeptide sequence refers to a polypeptide sequence that is shorter than the reference sequence but that retains essentially the same biological function or activity as the reference polypeptide. "Fragment" of a polynucleotide sequence refers to a polynucleotide sequence that is shorter than the reference sequence of SEQ ID NO:1..

"Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains the essential properties thereof. A typical variant of a polynucleotide differs in nucleotide sequence from the reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from the reference polypeptide. Generally, alterations are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, insertions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. Typical conservative substitutions include Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe and Tyr. A variant of a polynucleotide or polypeptide may be naturally occurring such as an allele, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis. Also included as variants are polypeptides having one or more post-translational modifications, for instance glycosylation, phosphorylation, methylation, ADP ribosylation and the like. Embodiments include methylation of the N-terminal amino acid, phosphorylations of serines and threonines and modification of C-terminal glycines.

"Allele" refers to one of two or more alternative forms of a gene occurring at a given locus in the genome.



"Polymorphism" refers to a variation in nucleotide sequence (and encoded polypeptide sequence, if relevant) at a given position in the genome within a population.

5 "Single Nucleotide Polymorphism" (SNP) refers to the occurrence of nucleotide variability at a single nucleotide position in the genome, within a population. An SNP may occur within a gene or within intergenic regions of the genome. SNPs can be assayed using Allele Specific Amplification (ASA). For the process at least 3 primers are required. A common primer is used in reverse complement to the polymorphism being assayed. This common primer can be between 50 and 1500 bps from the polymorphic base. The other two (or more) primers are identical to each other except that the final 3' base wobbles to match one of the two (or more) alleles that make up the polymorphism. Two (or more) PCR reactions are then conducted on sample DNA, each using the common primer and one of the Allele Specific Primers.

10 "Splice Variant" as used herein refers to cDNA molecules produced from RNA molecules initially transcribed from the same genomic DNA sequence but which have undergone alternative RNA splicing. Alternative RNA splicing occurs when a primary RNA transcript undergoes splicing, generally for the removal of introns, which results in the production of more than one mRNA molecule each of that may encode different amino acid sequences. The term splice variant also refers to the proteins encoded by the above cDNA molecules.

20 "Identity" reflects a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, determined by comparing the sequences. In general, identity refers to an exact nucleotide to nucleotide or amino acid to amino acid correspondence of the two polynucleotide or two polypeptide sequences, respectively, over the length of the sequences being compared.

25 "% Identity" - For sequences where there is not an exact correspondence, a "% identity" may be determined. In general, the two sequences to be compared are aligned to give a maximum correlation between the sequences. This may include inserting "gaps" in either one or both sequences, to enhance the degree of alignment. A % identity may be determined over the whole length of each of the sequences being

30

35

compared (so-called global alignment), that is particularly suitable for sequences of the same or very similar length, or over shorter, defined lengths (so-called local alignment), that is more suitable for sequences of unequal length.

5 "Similarity" is a further, more sophisticated measure of the relationship between two polypeptide sequences. In general, "similarity" means a comparison between the amino acids of two polypeptide chains, on a residue by residue basis, taking into account not only exact correspondences between a between pairs of residues, one from each of  
10 the sequences being compared (as for identity) but also, where there is not an exact correspondence, whether, on an evolutionary basis, one residue is a likely substitute for the other. This likelihood has an associated "score" from which the "% similarity" of the two sequences can then be determined.

15 Methods for comparing the identity and similarity of two or more sequences are well known in the art. Thus for instance, programs available in the Wisconsin Sequence Analysis Package, version 9.1 (Devereux J et al, Nucleic Acids Res, 12, 387-395, 1984, available from Genetics Computer Group, Madison, Wisconsin, USA), for example the  
20 programs BESTFIT and GAP, may be used to determine the % identity between two polynucleotides and the % identity and the % similarity between two polypeptide sequences. BESTFIT uses the "local homology" algorithm of Smith and Waterman (J Mol Biol, 147,195-197, 1981, Advances in Applied Mathematics, 2, 482-489, 1981) and finds the  
25 best single region of similarity between two sequences. BESTFIT is more suited to comparing two polynucleotide or two polypeptide sequences that are dissimilar in length, the program assuming that the shorter sequence represents a portion of the longer. In comparison, GAP aligns two sequences, finding a "maximum similarity", according to the  
30 algorithm of Neddleman and Wunsch (J Mol Biol, 48, 443-453, 1970). GAP is more suited to comparing sequences that are approximately the same length and an alignment is expected over the entire length. Preferably, the parameters "Gap Weight" and "Length Weight" used in  
35 each program are 50 and 3, for polynucleotide sequences and 12 and 4 for polypeptide sequences, respectively. Preferably, % identities and similarities are determined when the two sequences being compared are optimally aligned.

Other programs for determining identity and/or similarity between sequences are also known in the art, for instance the BLAST family of programs (Altschul S F et al, J Mol Biol, 215, 403-410, 1990, Altschul S F et al, Nucleic Acids Res., 25:389-3402, 1997, available from the National Center for Biotechnology Information (NCBI), Bethesda, Maryland, USA and accessible through the home page of the NCBI at [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) and FASTA (Pearson W R, Methods in Enzymology, 183, 63-99, 1990; Pearson W R and Lipman D J, Proc Nat Acad Sci USA, 85, 2444-2448, 1988, available as part of the Wisconsin Sequence Analysis Package).

Preferably, the BLOSUM62 amino acid substitution matrix (Henikoff S and Henikoff J G, Proc. Nat. Acad Sci. USA, 89, 10915-10919, 1992) is used in polypeptide sequence comparisons including where nucleotide sequences are first translated into amino acid sequences before comparison.

Preferably, the program BESTFIT is used to determine the % identity of a query polynucleotide or a polypeptide sequence with respect to a reference polynucleotide or a polypeptide sequence, the query and the reference sequence being optimally aligned and the parameters of the program set at the default value, as hereinbefore described.

"Identity Index" is a measure of sequence relatedness which may be used to compare a candidate sequence (polynucleotide or polypeptide) and a reference sequence. Thus, for instance, a candidate polynucleotide sequence having, for example, an Identity Index of 0.95 compared to a reference polynucleotide sequence is identical to the reference sequence except that the candidate polynucleotide sequence may include on average up to five differences per each 100 nucleotides of the reference sequence. Such differences are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion. These differences may occur at the 5' or 3' terminal positions of the reference polynucleotide sequence or anywhere between these terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. In other words, to obtain a polynucleotide sequence having an Identity Index of 0.95 compared to a reference polynucleotide sequence, an average of up to 5

in every 100 of the nucleotides of the in the reference sequence may be deleted, substituted or inserted, or any combination thereof, as hereinbefore described. The same applies *mutatis mutandis* for other values of the Identity Index, for instance 0.96, 0.97, 0.98 and 0.99.

5 Similarly, for a polypeptide, a candidate polypeptide sequence having, for example, an Identity Index of 0.95 compared to a reference polypeptide sequence is identical to the reference sequence except that the polypeptide sequence may include an average of up to five differences per each 100 amino acids of the reference sequence. Such differences  
10 are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion. These differences may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between these terminal positions, interspersed either  
15 individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. In other words, to obtain a polypeptide sequence having an Identity Index of 0.95 compared to a reference polypeptide sequence, an average of up to 5 in every 100 of the amino acids in the reference sequence may be deleted,  
20 substituted or inserted, or any combination thereof, as hereinbefore described. The same applies *mutatis mutandis* for other values of the Identity Index, for instance 0.96, 0.97, 0.98 and 0.99.

The relationship between the number of nucleotide or amino acid differences and the Identity Index may be expressed in the following  
25 equation:

$$n_a \leq x_a - (x_a \bullet I),$$

in which:

$n_a$  is the number of nucleotide or amino acid differences,

$x_a$  is the total number of nucleotides or amino acids in SEQ ID NO:1 or  
30 SEQ ID NO:2, respectively,

$I$  is the Identity Index ,

$\bullet$  is the symbol for the multiplication operator, and

in which any non-integer product of  $x_a$  and  $l$  is rounded down to the nearest integer prior to subtracting it from  $x_a$ .

"Homolog" is a generic term used in the art to indicate a polynucleotide or polypeptide sequence possessing a high degree of sequence relatedness to a reference sequence. Such relatedness may be quantified by determining the degree of identity and/or similarity between the two sequences as hereinbefore defined. Falling within this generic term are the terms "ortholog", and "paralog". "Ortholog" refers to a polynucleotide or polypeptide that is the functional equivalent of the polynucleotide or polypeptide in another species. "Paralog" refers to a polynucleotide or polypeptide that within the same species which is functionally similar.

"Fusion protein" refers to a protein encoded by two, unrelated, fused genes or fragments thereof. Examples have been disclosed in US 5541087, 5726044. In the case of Fc-DUSP10, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for performing the functional expression of Fc-DUSP10 or fragments of DUSP10, to improve pharmacokinetic properties of such a fusion protein when used for therapy and to generate a dimeric DUSP10. The Fc-DUSP10 DNA construct comprises in 5' to 3' direction, a secretion cassette, i.e. a signal sequence that triggers export from a mammalian cell, DNA encoding an immunoglobulin Fc region fragment, as a fusion partner, and a DNA encoding DUSP10 or fragments thereof. In some uses it would be desirable to be able to alter the intrinsic functional properties (complement binding, Fc-Receptor binding) by mutating the functional Fc sides while leaving the rest of the fusion protein untouched or delete the Fc part completely after expression.

All publications and references, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference in their entirety as if each individual publication or reference were specifically and individually indicated to be incorporated by reference herein as being fully set forth. Any patent application to which this application claims priority is also incorporated by reference herein in its entirety in the manner described above for publications and references.

### Further Examples

5 The partial Sequence identity among the EST AI424957 (see SEQ ID NO:3) and DUSP-10 has been investigated in more detail. This specific EST-sequence was derived from a cDNA library made of B-cells from a chronic lymphatic leukemia patient. It could be shown that the match among DUSP-10 (position 205-421) and AI424957 (position 205-421) was perfect, while there was a sudden brake in this homology outside that region, where the EST sequence did not indicate homology any longer. 10 This indicates that unexpected genomic aberrations within the DUSP-10 gene do occur and that in addition to the phosphatase protein function other functional aspects related to rearrangement of this gene do exist. Gene rearrangement in CLL may be of value to clinicians to know, because this has prognostic significance and does indicate the relevance of DUSP-10 as a novel tumor suppressor gene. 15

**Claims**

1. A polypeptide selected from the group consisting of:
  - (a) a polypeptide encoded by a polynucleotide comprising the sequence of SEQ  
5 ID NO:1;
  - (b) a polypeptide comprising a polypeptide sequence having at least 95% identity  
to the polypeptide sequence of SEQ ID NO:2;
  - (c) a polypeptide having at least 95% identity to the polypeptide sequence of  
SEQ ID NO:2;
  - 10 d) the polypeptide sequence of SEQ ID NO:2 and
  - (e) fragments and variants of such polypeptides in (a) to (d).
2. The polypeptide of claim 1 comprising the polypeptide sequence of SEQ ID  
NO:2.
- 15 3. The polypeptide of claim 1 which is the polypeptide sequence of SEQ ID  
NO:2.
4. A polynucleotide selected from the group consisting of:
  - (a) a polynucleotide comprising a polynucleotide sequence having at least 95%  
20 identity to the polynucleotide sequence of SEQ ID NO:1;
  - (b) a polynucleotide having at least 95% identity to the polynucleotide of SEQ ID  
NO:1;
  - (c) a polynucleotide comprising a polynucleotide sequence encoding a polypeptide  
sequence having at least 95% identity to the polypeptide sequence of SEQ ID  
25 NO:2;
  - (d) a polynucleotide having a polynucleotide sequence encoding a polypeptide  
sequence having at least 95% identity to the polypeptide sequence of SEQ ID  
NO:2;

- (e) a polynucleotide with a nucleotide sequence of at least 100 nucleotides obtained by screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof having at least 15 nucleotides;
- 5 (f) a polynucleotide which is the RNA equivalent of a polynucleotide of (a) to (e);
- (g) a polynucleotide sequence complementary to said polynucleotide of any one of (a) to (f), and
- (h) polynucleotides that are variants or fragments of the polynucleotides of any one of (a) to (g) or that are complementary to above mentioned polynucleotides,
- 10 over the entire length thereof.

5. A polynucleotide of claim 4 selected from the group consisting of:

- (a) a polynucleotide comprising the polynucleotide of SEQ ID NO:1;
- (b) the polynucleotide of SEQ ID NO:1;
- 15 (c) a polynucleotide comprising a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2; and
- (d) a polynucleotide encoding the polypeptide of SEQ ID NO:2.

6. An expression system comprising a polynucleotide capable of producing a

20 polypeptide of any one of claim 1-3 when said expression vector is present in a compatible host cell.

7. A recombinant host cell comprising the expression vector of claim 6 or a membrane thereof expressing the polypeptide of any one of claim 1-3.

25

8. A process for producing a polypeptide of any one of claim 1-3 comprising the step of culturing a host cell as defined in claim 7 under conditions sufficient for



the production of said polypeptide and recovering the polypeptide from the culture medium.

9. A fusion protein consisting of the Immunoglobulin Fc-region and a polypeptide  
5 any one one of claims 1-3.
10. An antibody immunospecific for the polypeptide of any one of claims 1 to 3.
11. A method for screening to identify compounds that stimulate or inhibit the function or level of the polypeptide of any one of claim 1-3 comprising a method selected from the group consisting of:
- 10 (a) measuring or, detecting, quantitatively or qualitatively, the binding of a candidate compound to the polypeptide (or to the cells or membranes expressing the polypeptide) or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound;
- (b) measuring the competition of binding of a candidate compound to the  
15 polypeptide (or to the cells or membranes expressing the polypeptide) or a fusion protein thereof in the presence of a labeled competitor;
- (c) testing whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells or cell membranes expressing the polypeptide;
- 20 (d) mixing a candidate compound with a solution containing a polypeptide of any one of claims 1-3, to form a mixture, measuring activity of the polypeptide in the mixture, and comparing the activity of the mixture to a control mixture which contains no candidate compound; or
- (e) detecting the effect of a candidate compound on the production of mRNA  
25 encoding said polypeptide or said polypeptide in cells, using for instance, an ELISA assay, and
- (f) producing said compound according to biotechnological or chemical standard techniques.

## SEQUENCE LISTING

<110> Merck Patent GmbH  
 5 <120> New dual specificity phosphatase  
 <130> DUSP10KDWS  
 10 <140>  
 <141>  
 <160> 2  
 15 <170> PatentIn Ver. 2.1  
 <210> 1  
 <211> 3059  
 <212> DNA  
 20 <213> Homo sapiens  
 <220>  
 <221> CDS  
 <222> (127)..(2121)  
 25 <400> 1  
 cagtcaccac cagcctgacc tcatacactt ttagtacaat ggagtggctg agcctttgag 60  
 cacaccacca ttacatcatc gtggcaaatt aaagaaggag gtgggaaaag aggacttatt 120  
 30 gttgtc atg gcc cat gag atg att gga act caa att gtt act gag agg 168  
 Met Ala His Glu Met Ile Gly Thr Gln Ile Val Thr Glu Arg  
 1 5 10  
 35 ttg gtg gct ctg ctg gaa agt gga acg gaa aaa gtg ctg cta att gat 216  
 Leu Val Ala Leu Leu Glu Ser Gly Thr Glu Lys Val Leu Leu Ile Asp  
 15 20 25 30  
 40 agc cgg cca ttt gtg gaa tac aat aca tcc cac att ttg gaa gcc att 264  
 Ser Arg Pro Phe Val Glu Tyr Asn Thr Ser His Ile Leu Glu Ala Ile  
 35 40 45  
 45 aat atc aac tgc tcc aag ctt atg aag cga agg ttg caa cag gac aaa 312  
 Asn Ile Asn Cys Ser Lys Leu Met Lys Arg Arg Leu Gln Gln Asp Lys  
 50 55 60  
 gtg tta att aca gag ctc atc cag cat tca gcg aaa cat aag gtt gac 360  
 Val Leu Ile Thr Glu Leu Ile Gln His Ser Ala Lys His Lys Val Asp  
 65 70 75  
 50 att gat tgc agt cag aag gtt gta gtt tac gat caa agc tcc caa gat 408  
 Ile Asp Cys Ser Gln Lys Val Val Val Tyr Asp Gln Ser Ser Gln Asp  
 80 85 90  
 55 gtt gcc tct ctc tct tca gac tgt ttt ctc act gta ctt ctg ggt aaa 456  
 Val Ala Ser Leu S r Ser Asp Cys Phe L u Thr Val Leu Leu Gly Lys  
 95 100 105 110  
 ctg gag aag agc ttc aac tct gtt cac ctg ctt gca ggt ggg ttt gct 504

	Leu	Glu	Lys	Ser	Phe	Asn	Ser	Val	His	Leu	Leu	Ala	Gly	Gly	Phe	Ala	
					115					120					125		
5	gag	ttc	tct	cgt	tgt	ttc	cct	ggc	ctc	tgt	gaa	gga	aaa	tcc	act	cta	552
	Glu	Phe	Ser	Arg	Cys	Phe	Pro	Gly	Leu	Cys	Glu	Gly	Lys	Ser	Thr	Leu	
				130					135					140			
10	gtc	cct	acc	tgc	att	tct	cag	cct	tgc	tta	cct	gtt	gcc	aac	att	ggg	600
	Val	Pro	Thr	Cys	Ile	Ser	Gln	Pro	Cys	Leu	Pro	Val	Ala	Asn	Ile	Gly	
				145					150					155			
15	cca	acc	cga	att	ctt	ccc	aat	ctt	tat	ctt	ggc	tgc	cag	cga	gat	gtc	645
	Pro	Thr	Arg	Ile	Leu	Pro	Asn	Leu	Tyr	Leu	Gly	Cys	Gln	Arg	Asp	Val	
				160				165				170					
20	ctc	aac	aag	gag	ctg	atg	cag	cag	aat	ggg	att	ggt	tat	gtg	tta	aat	696
	Leu	Asn	Lys	Glu	Leu	Met	Gln	Gln	Asn	Gly	Ile	Gly	Tyr	Val	Leu	Asn	
						180					185					190	
25	gcc	agc	aat	acc	tgt	cca	aag	cct	gac	ttt	atc	ccc	gag	tct	cat	ttc	744
	Ala	Ser	Asn	Thr	Cys	Pro	Lys	Pro	Asp	Phe	Ile	Pro	Glu	Ser	His	Phe	
					195					200					205		
30	ctg	cgt	gtg	cct	gtg	aat	gac	agc	ttt	tgt	gag	aaa	att	ttg	ccg	tgg	792
	Leu	Arg	Val	Pro	Val	Asn	Asp	Ser	Phe	Cys	Glu	Lys	Ile	Leu	Pro	Trp	
				210					215					220			
35	ttg	gac	aaa	tca	gta	gat	ttc	att	gag	aaa	gca	aaa	gcc	tcc	aat	gga	840
	Leu	Asp	Lys	Ser	Val	Asp	Phe	Ile	Glu	Lys	Ala	Lys	Ala	Ser	Asn	Gly	
				225				230					235				
40	tgt	gtt	cta	gtg	cac	tgt	tta	gct	ggg	atc	tcc	cgc	tcc	gcc	acc	atc	888
	Cys	Val	Leu	Val	His	Cys	Leu	Ala	Gly	Ile	Ser	Arg	Ser	Ala	Thr	Ile	
				240			245					250					
45	gct	atc	gcc	tac	atc	atg	aag	agg	atg	gac	atg	tct	tta	gat	gaa	gct	936
	Ala	Ile	Ala	Tyr	Ile	Met	Lys	Arg	Met	Asp	Met	Ser	Leu	Asp	Glu	Ala	
				255			260				265				270		
50	tac	aga	ttt	gtg	aaa	gaa	aaa	aga	cct	act	ata	tct	cca	aac	ttc	aat	984
	Tyr	Arg	Phe	Val	Lys	Glu	Lys	Arg	Pro	Thr	Ile	Ser	Pro	Asn	Phe	Asn	
				275					280						285		
55	ttt	ctg	ggc	caa	ctc	ctg	gac	tat	gag	aag	aag	att	aag	aac	cag	act	1032
	Phe	Leu	Gly	Gln	Leu	Leu	Asp	Tyr	Glu	Lys	Lys	Ile	Lys	Asn	Gln	Thr	
				290				295						300			
60	gga	gca	tca	ggg	cca	aag	agc	aaa	ctc	aag	ctg	ctg	cac	ctg	gag	aag	1080
	Gly	Ala	Ser	Gly	Pro	Lys	Ser	Lys	Leu	Lys	Leu	Leu	His	Leu	Glu	Lys	
				305				310					315				
65	cca	aat	gaa	cct	gtc	cct	gct	gtc	tca	gag	ggt	gga	cag	aaa	agc	gag	1128
	Pro	Asn	Glu	Pro	Val	Pro	Ala	Val	Ser	Glu	Gly	Gly	Gln	Lys	Ser	Glu	
				320			325					330					
70	acg	ccc	ctc	agt	cca	ccc	tgt	gcc	gac	tct	gct	acc	tca	gag	gca	gca	1176
	Thr	Pro	Lys	Ser	Pro	Pro	Cys	Ala	Asp	Ser	Ala	Thr	Ser	Glu	Ala	Ala	
				335			340				345				350		
75	gga	caa	agg	ccc	gtg	cat	ccc	gcc	agc	gtg	ccc	agc	gtg	ccc	agc	gtg	1224
	Gly	Gln	Arg	Pro	Val	His	Pro	Ala	Ser	Val	Pro	Ser	Val	Pro	Ser	Val	

	355	360	365	
5	cag ccg tcg ctg tta gag gac agc ccg ctg gta cag gcg ctc agt ggg Gln Pro Ser Leu Leu Glu Asp Ser Pro L u Val Gln Ala Leu Ser Gly 370 375 380	1272		
10	ctg cac ctg tcc gca gac agg ctg gaa gac agc aat aag ctc aag cgt Leu His Leu Ser Ala Asp Arg Leu Glu Asp Ser Asn Lys Leu Lys Arg 385 390 395	1320		
15	tgc ttc tct ctg gat atc aaa tca gtt tca tat tca gcc agc atg gca Ser Phe Ser Leu Asp Ile Lys Ser Val Ser Tyr Ser Ala Ser Met Ala 400 405 410	1368		
20	gca tcc tta cat ggc ttc tcc tca tca gaa gat gct ttg gaa tac tac Ala Ser Leu His Gly Phe Ser Ser Ser Glu Asp Ala Leu Glu Tyr Tyr 415 420 425 430	1416		
25	aaa cct tcc act act ctg gat ggg acc aac aag cta tgc cag ttc tcc Lys Pro Ser Thr Thr Leu Asp Gly Thr Asn Lys Leu Cys Gln Phe Ser 435 440 445	1464		
30	cct gtt cag gaa cta tcg gag cag act ccc gaa acc agt cct gat aag Pro Val Gln Glu Leu Ser Glu Gln Thr Pro Glu Thr Ser Pro Asp Lys 450 455 460	1512		
35	gag gaa gcc agc atc ccc aag aag ctg cag acc gcc agg cct tca gac Glu Glu Ala Ser Ile Pro Lys Lys Leu Gln Thr Ala Arg Pro Ser Asp 465 470 475	1560		
40	agc cag agc aag cga ttg cat tcg gtc aga acc agc agc agt ggc acc Ser Gln Ser Lys Arg Leu His Ser Val Arg Thr Ser Ser Ser Gly Thr 480 485 490	1608		
45	gcc cag agg tcc ctt tta tct cca ctg cat cga agt ggg agc gtg gag Ala Gln Arg Ser Leu Leu Ser Pro Leu His Arg Ser Gly Ser Val Glu 495 500 505 510	1656		
50	gac aat tac cac acc agc ttc ctt ttc ggc ctt tcc acc agc cag cag Asp Asn Tyr His Thr Ser Phe Leu Phe Gly Leu Ser Thr Ser Gln Gln 515 520 525	1704		
55	cac ctc acg aag tct gct ggc ctg ggc ctt aag ggc tgg cac tcg gat His Leu Thr Lys Ser Ala Gly Leu Gly Leu Lys Gly Trp His Ser Asp 530 535 540	1752		
60	atc ttg gcc ccc cag acc tct acc cct tcc ctg acc agc agc tgg tat Ile Leu Ala Pro Gln Thr Ser Thr Pro Ser Leu Thr Ser Ser Trp Tyr 545 550 555	1800		
65	ttt gcc aca gag tcc tca cac ttc tac tct gcc tca gcc atc tac gga Phe Ala Thr Glu Ser Ser His Phe Tyr Ser Ala Ser Ala Ile Tyr Gly 560 565 570	1848		
70	ggc agt gcc agt tac tct gcc tac agc tgc agc cag ctg ccc act tgc Gly Ser Ala Ser Tyr Ser Ala Tyr Ser Cys Ser Gln Leu Pro Thr Cys 575 580 585 590	1896		
75	gga gac caa gtc tat tct gtg cgc agg cgg cag aag cca agt gac aga Gly Asp Gln Val Tyr Ser Val Arg Arg Gln Lys Pro Ser Asp Arg 595 600 605	1944		

gct gac tcg cgg cgg agc tgg cat gaa gag agc ccc ttt gaa aag cag 1992  
 Ala Asp Ser Arg Arg Ser Trp His Glu Glu Ser Pro Phe Glu Lys Gln  
 610 615 620

5 ttt aaa cgc aga agc tgc caa atg gaa ttt gga gag agc atc atg tca 2040  
 Phe Lys Arg Arg Ser Cys Gln Met Glu Phe Gly Glu Ser Ile Met Ser  
 625 630 635

10 gag aac agg tca cgg gaa gag ctg ggg aaa gtg ggc agt cag tct agc 2088  
 Glu Asn Arg Ser Arg Glu Glu Leu Gly Lys Val Gly Ser Gln Ser Ser  
 640 645 650

15 ttt tcg ggc agc atg gaa atc att gag gtc tcc tgagaagaaa gacacttggtg 2141  
 Phe Ser Gly Ser Met Glu Ile Ile Glu Val Ser  
 655 660 665

acttctatag acaattttttt tttcttggtc acaaaaaaat tccctgtaaa tctgaaatat 2201

20 atatatgtac atacatatat attttttgaa aatggagcta tgggtgtaaaa gcaacagggtg 2261  
 gatcaaccca gttgttactc tcttaacatc tgcatttgag agatcagcta atacttctct 2321  
 caacaaaaat ggaagggcag atgctagaat cccccctaga cggaggaaaa ccattttatt 2381

25 cagtgaatta cacatcctct tgttcttaaa aaagcaagtg tctttggtgt tggaggacaa 2441  
 aatcccctac cattttcacg ttgtgctact aagagatctc aaatattagt ctttgtccgg 2501

30 acccttccat agtacacctt agcgttgaga ctgagccagc ttgggggtca ggtaggtaga 2561  
 ccttgtagg gacagagcct agtggtaaat ccaagagaaa tgatcctatc caaagctgat 2621  
 tcacaaaccc acgctcacct gacagccgag ggacacgagc atcactctgc tggacggacc 2681

35 attaggggccc ttgccaaggt ctaccttaga gcaaaccag tacctcagac aggaaagtcg 2741  
 gggctttgac cactaccata tctggtagcc cattttctag gcattgtgaa taggtaggta 2801

40 gctagtcaca cttttcagac caattcaaac tgtctatgca caaaattccc gtgggcctag 2861  
 atggagataa ttttttttct ttctcagctt tatgaagaga agggaaactg tctaggattc 2921  
 agctgaacca ccaggaacct ggcaacatca cgatttaagc taagggttggg aggctaacga 2981

45 gtctacctcc ctctttgtaa atcaaagaat tgtttaaaat gggattgtca atcctttaaa 3041  
 taaagatgaa cttgggtt 3059

50 <210> 2  
 <211> 665  
 <212> PRT  
 <213> Homo sapiens

55 <400> 2  
 Met Ala His Glu Met Ile Gly Thr Gln Ile Val Thr Glu Arg Leu Val  
 1 5 10 15

60 Ala Leu Leu Glu Ser Gly Thr Glu Lys Val Leu Leu Ile Asp Ser Arg  
 20 25 30

Pro Phe Val Glu Tyr Asn Thr Ser His Ile Leu Glu Ala Ile Asn Ile  
                   35                                  40                                  45

5 Asn Cys Ser Lys Leu Met Lys Arg Arg Leu Gln Gln Asp Lys Val Leu  
                   50                                  55                                  60

Ile Thr Glu Leu Ile Gln His Ser Ala Lys His Lys Val Asp Ile Asp  
           65                                  70                                  75                                  80

10 Cys Ser Gln Lys Val Val Val Tyr Asp Gln Ser Ser Gln Asp Val Ala  
                                   85                                  90                                  95

15 Ser Leu Ser Ser Asp Cys Phe Leu Thr Val Leu Leu Gly Lys Leu Glu  
                                   100                                  105                                  110

Lys Ser Phe Asn Ser Val His Leu Leu Ala Gly Gly Phe Ala Glu Phe  
                   115                                  120                                  125

20 Ser Arg Cys Phe Pro Gly Leu Cys Glu Gly Lys Ser Thr Leu Val Pro  
                   130                                  135                                  140

Thr Cys Ile Ser Gln Pro Cys Leu Pro Val Ala Asn Ile Gly Pro Thr  
           145                                  150                                  155                                  160

25 Arg Ile Leu Pro Asn Leu Tyr Leu Gly Cys Gln Arg Asp Val Leu Asn  
                                   165                                  170                                  175

30 Lys Glu Leu Met Gln Gln Asn Gly Ile Gly Tyr Val Leu Asn Ala Ser  
                                   180                                  185                                  190

Asn Thr Cys Pro Lys Pro Asp Phe Ile Pro Glu Ser His Phe Leu Arg  
                   195                                  200                                  205

35 Val Pro Val Asn Asp Ser Phe Cys Glu Lys Ile Leu Pro Trp Leu Asp  
                   210                                  215                                  220

Lys Ser Val Asp Phe Ile Glu Lys Ala Lys Ala Ser Asn Gly Cys Val  
           225                                  230                                  235                                  240

40 Leu Val His Cys Leu Ala Gly Ile Ser Arg Ser Ala Thr Ile Ala Ile  
                                   245                                  250                                  255

45 Ala Tyr Ile Met Lys Arg Met Asp Met Ser Leu Asp Glu Ala Tyr Arg  
                                   260                                  265                                  270

Phe Val Lys Glu Lys Arg Pro Thr Ile Ser Pro Asn Phe Asn Phe Leu  
                   275                                  280                                  285

50 Gly Gln Leu Leu Asp Tyr Glu Lys Lys Ile Lys Asn Gln Thr Gly Ala  
                   290                                  295                                  300

Ser Gly Pro Lys Ser Lys Leu Lys Leu Leu His Leu Glu Lys Pro Asn  
           305                                  310                                  315                                  320

55 Glu Pro Val Pro Ala Val Ser Glu Gly Gly Gln Lys Ser Glu Thr Pro  
                                   325                                  330                                  335

60 Leu Ser Pro Pro Cys Ala Asp Ser Ala Thr Ser Glu Ala Ala Gly Gln  
                                   340                                  345                                  350

Arg Pro Val His Pro Ala Ser Val Pro Ser Val Pro Ser Val Gln Pro  
 355 360 365  
 5 Ser Leu Leu Glu Asp Ser Pro Leu Val Gln Ala Leu Ser Gly Leu His  
 370 375 380  
 Leu Ser Ala Asp Arg Leu Glu Asp Ser Asn Lys Leu Lys Arg Ser Phe  
 385 390 395 400  
 10 Ser Leu Asp Ile Lys Ser Val Ser Tyr Ser Ala Ser Met Ala Ala Ser  
 405 410 415  
 Leu His Gly Phe Ser Ser Ser Glu Asp Ala Leu Glu Tyr Tyr Lys Pro  
 420 425 430  
 15 Ser Thr Thr Leu Asp Gly Thr Asn Lys Leu Cys Gln Phe Ser Pro Val  
 435 440 445  
 Gln Glu Leu Ser Glu Gln Thr Pro Glu Thr Ser Pro Asp Lys Glu Glu  
 450 455 460  
 20 Ala Ser Ile Pro Lys Lys Leu Gln Thr Ala Arg Pro Ser Asp Ser Gln  
 465 470 475 480  
 25 Ser Lys Arg Leu His Ser Val Arg Thr Ser Ser Ser Gly Thr Ala Gln  
 485 490 495  
 Arg Ser Leu Leu Ser Pro Leu His Arg Ser Gly Ser Val Glu Asp Asn  
 500 505 510  
 30 Tyr His Thr Ser Phe Leu Phe Gly Leu Ser Thr Ser Gln Gln His Leu  
 515 520 525  
 Thr Lys Ser Ala Gly Leu Gly Leu Lys Gly Trp His Ser Asp Ile Leu  
 530 535 540  
 35 Ala Pro Gln Thr Ser Thr Pro Ser Leu Thr Ser Ser Trp Tyr Phe Ala  
 545 550 555 560  
 40 Thr Glu Ser Ser His Phe Tyr Ser Ala Ser Ala Ile Tyr Gly Gly Ser  
 565 570 575  
 Ala Ser Tyr Ser Ala Tyr Ser Cys Ser Gln Leu Pro Thr Cys Gly Asp  
 580 585 590  
 45 Gln Val Tyr Ser Val Arg Arg Arg Gln Lys Pro Ser Asp Arg Ala Asp  
 595 600 605  
 Ser Arg Arg Ser Trp His Glu Glu Ser Pro Phe Glu Lys Gln Phe Lys  
 610 615 620  
 50 Arg Arg Ser Cys Gln Met Glu Phe Gly Glu Ser Ile Met Ser Glu Asn  
 625 630 635 640  
 55 Arg Ser Arg Glu Glu Leu Gly Lys Val Gly Ser Gln Ser Ser Phe Ser  
 645 650 655  
 Gly Ser Met Glu Ile Ile Glu Val Ser  
 660 665  
 60

<210> 3  
<211> 246  
<212> DNA  
<213> Homo sapiens  
5 <400> 3  
ctgctaattg atagccggcc atttgtggaa tacaatacat cccacatttt ggaagccatt 60  
aatatcaact gtcceaagct tatgaagcga aggttgcaac aggacaaagt gttaattaca 120  
10 gagctcatcc agcattcagc gaaacataag gttgacattg attgcagtca gaagggttga 180  
gtttacgata aaagctccca agatgttgcc tctctcttca gcatgtaatt ggggaaaccc 240  
tcgttt 246  
15



# INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 01/03966

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/55 C12N15/62 C12N9/16 C07K16/40 C12Q1/68  
C12Q1/42 G01N33/53

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K C12Q G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL 'Online! Entry/Acc.no. AC007619, 24 May 1999 (1999-05-24) MUNZY, D.M. ET AL.: "Homo sapiens 12p BAC RP11-253I19 (Roswell Park Cancer Institute Human BAC" XP002176719 nt. 101430-103560.	4
X	DATABASE EMBL 'Online! Entry/Acc.no. AI659120, 11 May 1999 (1999-05-11) STRAUSBERG, R.: "tu08d11.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2250453 3', mRNA sequence." XP002176720 the whole document	4



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

\* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*8\* document member of the same patent family

Date of the actual completion of the international search

5 September 2001

Date of mailing of the international search report

19/09/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Smalt, R

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 01/03966

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	NAGASE T, KIKUNO R, HATTORI A, KONDO Y, OKUMURA K, OHARA O.: "Prediction of the coding sequences of unidentified human genes. XIX. The complete" DNA RESEARCH, vol. 7, no. 6, 31 December 2000 (2000-12-31), pages 347-55, XP001016382 see KIAA1700	1,2,4, 6-8
P, X	--- DATABASE EMBL 'Online! Entry/Acc.no. AB052156, 5 April 2001 (2001-04-05) MASUDA, K. ET AL.: "Homo sapiens MKP-7 mRNA for MAPK phosphatase-7, complete cds." XP002176721 the whole document	1-8
P, X	--- WO 01 20004 A (INCYTE GENOMICS INC ;AZIMZAI YALDA (US); YUE HENRY (US); BANDMAN O) 22 March 2001 (2001-03-22) see seq.ID's 4 and 15	1,4,6-11
E	--- WO 01 46394 A (PLOWMAN GREGORY D ;HILL RONALD J (US); SUGEN INC (US); WHYTE DAVID) 28 June 2001 (2001-06-28) claims, seq.ID's 2 and 14 (SGP002) page 123 -----	1-11

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 01/03966

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0120004 A	22-03-2001	AU 7829700 A	17-04-2001
WO 0146394 A	28-06-2001	WO 0155356 A	02-08-2001